

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
ІМЕНІ ІГОРЯ СІКОРСЬКОГО»**

ШАДУРА ОКСАНА ВІТАЛІЇВНА



УДК 004.942; 519.876.5

**СТОХАСТИЧНА ОПТИМІЗАЦІЯ
ПРОДУКТИВНОСТІ ПАКЕТУ СИМУЛЯЦІЇ ТРАНСПОРТУ
ЧАСТИНОК В ФІЗИЦІ ВИСОКИХ ЕНЕРГІЙ**

05.13.12 – Системи автоматизації проектувальних робіт

Автореферат дисертації на здобуття наукового ступеня
кандидата технічних наук

Київ – 2019

Дисертацією є рукопис.

Робота виконана на кафедрі системного проектування Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» Міністерства освіти і науки України.

Науковий керівник доктор технічних наук, професор
Петренко Анатолій Іванович
Національний технічний університет України
"Київський політехнічний інститут імені Ігоря Сікорського", завідувач кафедри системного проектування

Офіційні опоненти доктор технічних наук, професор
Безкоровайний Володимир Валентинович
Харківський національний університет
радіоелектроніки, професор кафедри системотехніки

доктор технічних наук, старший науковий співробітник
Чемерис Олександр Анатолійович
Інститут проблем моделювання в енергетиці
ім. Г.Є. Пухова НАН України, заступник директора з наукової роботи

Захист відбудеться 9 грудня 2019 р. о 14 годині 00 хвилин на засіданні спеціалізованої вченої ради К 26.002.17 у Національному технічному університеті України "Київський політехнічний інститут імені Ігоря Сікорського" за адресою: 03056, м.Київ, проспект Перемоги, 37, корп. 35, ауд. 120.

З дисертацією можна ознайомитись у науково-технічній бібліотеці Національного технічного університету України "Київський політехнічний інститут імені Ігоря Сікорського" за адресою: 03056, м. Київ, пр. Перемоги, 37.

Автореферат розіслано «___» _____ 2019 р.

Учений секретар
Спеціалізованої вченої ради К 26.002.17
кандидат технічних наук, с.н.с.

Г.Д. Кисельов

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми дисертації. На жаль, існує протиріччя в методах оптимізації продуктивності обчислень у САЕ-системах автоматизованого проектування (CAD). Для його розв'язання була сформульована наукова-прикладна задача, що дозволяє дослідити оптимізацію продуктивності програмних додатків для САЕ-системи автоматизованого проектування (CAD) в фізиці високих енергій, використовуючи алгоритми стохастичної оптимізації на прикладі оптимізації продуктивності пакету моделювання (симуляцій) транспорту частинок GeantV.

GeantV є прототипом програмного забезпечення наступного покоління для моделювання транспорту частинок через різні типи матеріалів і має увібрати в себе та розвинути всі позитивні сторони Geant4 САЕ-системи автоматизованого проектування і розрахунків, яка безумовно є стандартом для обробки даних у фізиці високих енергій (ФВЕ), ядерній фізиці, фізиці прискорювачів, медицині та фізиці космосу, і яку, зокрема використовують для аналізу даних в експериментах на Великому адронному колайдері (ВАК), розташованому в ЦЕРН, м. Женева, Швейцарія.

Не дивлячись на багаточисленні публікації по цій темі, в результаті її аналізу було знайдено протиріччя між необхідністю збільшення ефективності існуючих підходів і обмеженість математичних моделей та методів розв'язання задач оптимізації систем проектування.

Основна проблематика полягає у тому, що експерименти на прискорювачі ВАК генерують велику кількість даних, обробка яких вимагає більш ефективного використання обчислювальної техніки нового покоління та потребує розробки нових підходів до методів розв'язання задач оптимізації систем проектування. Основними особливостями цих задач є:

- наявність великого та не гетерогенного пулу ресурсів, що задіяний для аналізу даних на прискорювачі ВАК та вимагає додаткових зусиль для оптимізації продуктивності і планування завдань проведення паралельних обчислень при обробленні даних;

- алгоритмізація завдань проектування за допомогою методів машинного навчання, які надають можливість ефективніше використовувати ресурси та проводити подальше прогнозування оптимальних параметрів для конкретного обчислювального середовища.

З урахуванням того що, для аналізу даних в експериментах на Великому адронному колайдері (ВАК) використовують грід-технології, які розподіляють обробку експериментальних даних серед великої кількості окремих обчислювальних кластерів, що побудовані з різних за

структурою комп'ютерних платформ, у цьому випадку пріоритет полягає у використанні контрольованого машинного навчання для класифікації подій для об'єкта реконструкції. У цьому напрямку вагомі результати отримали І.Фостер, К.Кессельман, Ф. Корбато, А.І. Петренко, С. Я. Свістунов, П.В. Свірін та інші.

Зв'язок роботи з науковими програмами, планами, темами.

Дисертацію виконано відповідно до плану науково-дослідної роботи кафедри системного проектування НТУУ «Київський політехнічний інститут імені Ігоря Сікорського» згідно з планами дослідження за темою: «Дослідження нової концепції побудови динамічної архітектури проблемно-орієнтованого програмного забезпечення в грід-хмарному середовищі з елементами постбінарних обчислень» (№ 2710-ф, номер державної реєстрації 0114U003449, терміни виконання: 2014-2016 рр.) та плану досліджень у рамках програми CERN Doctoral Student Programme 2014-2017, відділу SoFTware at Physics Department (EP-SFT), ЦЕРН, Женева, Швейцарія.

Мета та завдання дослідження. Метою дисертаційної роботи є розробка математичних моделей та методів оптимізації продуктивності САЕ-систем автоматизованого проектування для ФВЕ для збільшення продуктивності пакету моделювання транспорту частинок GeantV за допомогою оптимізації генетичних алгоритмів та їх програмної реалізації. Досягнення цієї мети полягає у вирішенні наступних задач:

- провести огляд і аналіз сучасного стану проблеми;
- розробити математичну модель для оптимізації роботи САЕ-системи моделювання транспорту частинок GeantV для обробки даних на ВАК на основі генетичних алгоритмів оптимізації багатокритеріальних задач;
- виконати формалізацію методу нецентрованих головних компонент та знаходження оцінки наближення при його застосуванні до аналізу даних;
- розробити моделі методу нецентрованих головних компонент та метод оцінки наближень при застосуванні методу для аналізу даних;
- розробити алгоритми та програмне забезпечення для вирішення завдань включення в генетичний алгоритм нового оператора генетичного алгоритму на основі методу нецентрованих головних компонент (НГК-оператор);
- перевірити ефективність його застосування для еталонних тестів багатокритеріальних задач;
- провести порівняльний аналіз ефективності включення НГК-оператора в генетичний алгоритм для оптимізації продуктивності САЕ-системи моделювання транспорту частинок GeantV в порівнянні з стандартними реалізаціями генетичних алгоритмів.

Об'єктом теоретичних і експериментальних досліджень відповідно до поставленої мети є САЕ-система моделювання транспорту частинок GeantV.

Предметом дослідження дисертаційної роботи є процес стохастичної оптимізації продуктивності САЕ-системи моделювання транспорту частинок GeantV на основі модифікації генетичних алгоритмів за допомогою методу нецентрованих головних компонент.

Методи досліджень. Для розв'язку задачі стохастичної оптимізації продуктивності САЕ-системи моделювання транспорту частинок GeantV в фізиці високих енергій використовувалися методи математичного моделювання, штучного інтелекту та інженерії програмного забезпечення, апарат теорії ймовірності та математичої статистики.

Методами досліджень є:

- генетичні алгоритми для оптимізації продуктивності САЕ-системи моделювання транспорту частинок GeantV;
- представлення генетичного алгоритму як динамічної системи на основі його зв'язку з марковськими ланцюжками;
- метод нецентрованих головних компонент;
- модифікація генетичного алгоритму за допомогою введення нового НГК-оператора в динамічну систему генетичного алгоритму.

Наукова новизна отриманих результатів.

Вперше:

- запропоновано математичну модель для оптимізації роботи САЕ-системи моделювання транспорту частинок GeantV для обробки даних у ВАК на основі генетичних алгоритмів для розв'язання типових багатокритеріальних задач;
- запропоновано математичну модель методу нецентрованих головних компонент та отримано оцінку наближень при застосуванні методу для аналізу даних, який характеризується тим, що пришвидчує перевірку чутливості параметрів та забезпечує збереження максимуму інформації при цих наближеннях;
- запропоновано метод прискорення збіжності генетичного алгоритму до оптимального Парето-фронту для розв'язання задач багатокритеріальної оптимізації;
- розроблено застосування запропонованого методу прискорення збіжності генетичного алгоритму до оптимального Парето-фронту на основі модифікації генетичних алгоритмів для оптимізації продуктивності програмового пакету САЕ-системи моделювання транспорту частинок GeantV в фізиці високих енергій.

Практичне значення отриманих результатів.

Практичне значення мають:

- розроблений в дисертаційній роботі оригінальний алгоритм для включення нового генетичного оператора (НГК-оператор), що надає змогу

прискорити збіжність генетичного алгоритму до оптимального Парето-фронту і тим самим прискорити отримання наближеного рішення для задач багатокритеріальної оптимізації;

- проведений порівняльний аналіз використання методу головних компонент при його застосуванні для центрованих і нецентрованих даних, та знайдена еквівалентність використання обох представлень даних на основі оцінки середньоквадратичної похибки;

- розроблена оптимальна інтеграція НГК-оператора в типовий генетичний алгоритм та досліджена ефективність його застосування на прикладі еталонних тестів для багатокритеріальних задач;

- розроблена в дисертаційній роботі бібліотека програмних модулів для застосування в САЕ-системі моделювання транспорту частинок GeantV, яка буде використовуватись для оптимізації продуктивності програмового пакету GeantV при його застосуванні для обробки даних в експериментах на Великому адронному колайдері.

Особистий внесок здобувача. Наукові положення, висновки і рекомендації, що виносяться на захист, отримано здобувачем самостійно. Із наукових праць, опублікованих у співавторстві, у дисертації використані лише ті ідеї, положення та матеріали, які є результатом власних досліджень автора. У роботах, написаних у співавторстві, автору належать наступні результати: у [1] розроблено математичний метод нецентрованих головних компонент (НГК), отримано оцінку наближень при застосуванні методу для аналізу даних та запропоновано модифікацію генетичних алгоритмів за допомогою НГК-оператора для оптимізації пакетів фізичного моделювання; у [2,3] запропоновано модифікацію генетичних алгоритмів за допомогою НГК-оператора для оптимізації пакетів фізичного моделювання; проведений аналіз вибору оптимальної модифікації генетичного алгоритму за рахунок включення в алгоритм НГК-оператора та досліджена продуктивність такого генетичного алгоритму на наборі стандартних тестів DTLZ для оптимізації багатокритеріальних задач; у [4, 5] сформульовано підхід до оптимізації САЕ-системи автоматизованого проектування GeantV, що спрямований на розробку спеціальних методів оптимізації продуктивності та планування завдань при проведенні паралельних обчислень при обробці даних з застосуванням цієї САЕ-системи; у [9, 10, 11, 12] досліджено вибір ключових параметрів САЕ-системи GeantV для оптимізації її продуктивності та перевірено ефективність цього набору параметрів при дослідженні застосування оптимізованої САЕ-системи GeantV в порівнянні з неоптимізованим пакетом на прикладі обробки експериментальних даних на різних процесорах Intel (R) Core i7 та Intel (R) Ivy Bridge; у [7, 8] показано, що модифікований генетичний алгоритм з включенням НГК-оператора прискорює оптимізацію продуктивності САЕ-системи GeantV при моделюванні транспорту частинок через простий

калориметр та збільшує ефективність використання обчислювальних ресурсів для аналізу даних у фізиці високих енергій; у [6] проведено аналіз слабких та сильних сторін CAE-системи автоматизованого проектування Geant4, яка є попередником програмового пакету GeantV, на основі аналізу даних в експерименті SHiP.

Апробація результатів дисертації. Основні положення та результати дисертаційної роботи доповідалися, обговорювалися та отримали схвальну оцінку на міжнародних конференціях та семінарах: IEEE Symposium Series on Computational Intelligence, SSCI 2016, (Афіни, Греція, 2016, 2016); Міжнародна конференція «Computing in High Energy Physics, CHER 2016» (Сан-Франциско, США, 2016); Міжнародна конференція «Computing in High Energy Physics, CHER 2015», (Окінава, Японія, 2016); «16-й Міжнародний семінар з передових обчислювальних та аналітичних методів у фізиці», ACAT 2014, (Прага, Чеська Республіка, 2014).

Публікації. За результатами досліджень опубліковано 12 наукових праць, у тому числі, 5 статей у наукових фахових виданнях (з них 2 статті у виданнях іноземних держав, 3 у виданнях України, 2 з яких включені до міжнародних наукометричних баз), 7 статей у збірниках праць міжнародних наукових конференцій.

Структура та об'єм роботи. Дисертаційна робота викладена на 208 сторінках машинописного тексту, складається зі вступу, 4 розділів, загальних висновків, списку використаних джерел та 4 додатка. Обсяг основного тексту дисертації складає 168 сторінок друкованого тексту. Робота ілюстрована 66 рисунками. Список використаних джерел містить 121 найменування, з них 3 найменування кирилицею та 118 найменувань латиницею.

ОСНОВНИЙ ЗМІСТ РОБОТИ

Протягом останнього десятиліття CAE-система Geant4 була основним програмним пакетом для повного моделювання детекторів у фізиці високих енергій. Проте через його складний дизайн, що має глибокі ієрархічні структури класів та стеків викликів, Geant4 важко експлуатувати на кращих сучасних комп'ютерних архітектурах, що значно проєволюціонували в останнє десятиліття. Вони надають потужні можливості для паралелізму і векторизації, зокрема роблять більш гнучким процес використання кешу інструкцій і кешу даних та підвищують локальність даних для використання в програмних додатках.

У 2013 році в CERN розпочався проєкт по розробці нової CAE-системи моделювання транспорту частинок GeantV, яка повинна вибрати в

себе та розвинути всі позитивні сторони пакету Geant4 та розширити функціональність, зокрема збільшити частку паралелізму і пропускну спроможність обробки даних за допомогою включення SIMD стратегій (англ. single instruction, multiple data) в код САЕ-системи GeantV.

У **Вступі** обґрунтовано актуальність теми дослідження, показано зв'язок проблеми з науковими програмами, планами та темами, сформульовано методи та основні завдання досліджень, наукову новизну та практичне значення отриманих результатів. Наведено дані про апробацію результатів роботи, публікації та особистий внесок здобувача.

У **першому розділі** *«Структурні елементи та алгоритми багатокритеріальної задачі оптимізації для пакету симуляції транспорту частинок GeantV»* проведено огляд стохастичних алгоритмів, обґрунтовано вибір генетичних алгоритмів для стохастичної оптимізації продуктивності пакету GeantV та вказано на їх переваги та недоліки. Визначено основні об'єкти та інструменти проведення такої оптимізації.

Метод генетичного алгоритму (ГА) є одним з широко використовуваних еволюційних методів для різних задач оптимізації і відноситься до методів стохастичної оптимізації. Стохастична оптимізація визначається, як знаходження глобального мінімуму або максимуму сукупності цільових функцій за наявності випадковості в процесі оптимізації. Для дослідження оптимізації роботи пакету GeantV застосовані генетичні алгоритми, методи машинного навчання, включаючи багатовимірний статистичний аналіз. Ці підходи придатні для розв'язування багатокритеріальної задачі «чорного ящика» з великими обсягами обчислень для оцінки функцій пристосованості. Гарним прикладом у цьому випадку є генетичні алгоритми з невідомим сортуванням з непрямыми зворотніми зв'язками (NSGA-II або NSGA-III алгоритми). Головним недоліком схеми генетичних алгоритмів є відсутність додаткових процедур перевірки або спеціальних операторів, що створюють механізми прискорення збіжності цих алгоритмів до множини глобальних оптимальних точок.

Теорія генетичних алгоритмів була предметом інтенсивних досліджень протягом останніх десятиліть. Базовою моделлю для генетичного алгоритму, що допускає точне математичне формулювання, є так звана «проста модель генетичного алгоритму» (ПГА), досліджена у працях М. Во́зе (1999). Ця модель описує генетичний алгоритм як динамічну систему і добре вивчена в літературі.

У цій моделі генетичний алгоритм розглядається як марковський ланцюжок, в якому стани еволюційної системи є популяціями, а переходи між ними керуються генетичними операторами, таким як селекція, мутація та схрещування (кросовер). Мутація гарантує, що ланцюжок Маркова є зв'язаним, тому існує унікальний рівноважний розподіл за популяціями. Ймовірність виникнення конкретної популяції в одному поколінні

залежить лише від факторів попереднього покоління та зовнішніх впливів. Цей стохастичний процес описується ланцюжком Маркова, що характеризується матрицею переходу $T_{\vec{q}, \vec{p}}$ від популяції \vec{p} до популяції \vec{q} , де у векторі популяцій

$$\left\{ \vec{p} = (p_1, p_2, \dots, p_m)^t, \quad 0 \leq p_\alpha \leq 1, \sum_{\alpha=1}^m p_\alpha = 1 \right\},$$

компонента p_α є ймовірністю появи в популяції α -ої особини. Ми розглядаємо популяції з m різними типами індивідів у просторі елементарних подій (простір індивідів) Ω .

Генетичний алгоритм, як динамічна система, описує еволюцію індивідів у просторі Ω для можливих популяцій фіксованого розміру m , де m можна розглядати як кількість вимірювань в ході експерименту.

Генетичні оператори (селекція, мутація та схрещування) діють у векторному просторі $\Lambda = \{(p_1, p_2, \dots, p_m)^t\}$ векторів популяцій. Генетичний оператор $G_\alpha(\vec{p})$ задає певну ймовірність отримання індивіда α в наступному поколінні, якщо попереднє покоління фіксується вектором \vec{p} , і повний генетичний оператор визначається відображенням $G: \Lambda \rightarrow \Lambda$, де $G(\vec{p}) = \prod_{\alpha \in \Omega} G_\alpha(\vec{p})$, є композицією операторів: селекції, мутації і кросовера.

Визначимо оператор генетичного відбору (селекції) $F: \Lambda \rightarrow \Lambda$, де $F(\vec{p}) = \prod_{\alpha \in \Omega} F_\alpha(\vec{p})$ і α -та компонента $F_\alpha(\vec{p})$ є ймовірністю появи індивіда типу α , якщо селекцію застосовують до $\vec{p} \in \Lambda$.

Оператор мутації $\hat{U}: \Lambda \rightarrow \Lambda$ є матрицею розміру $m \times m$ з дійсними матричними елементами $u_{\alpha, \beta} > 0$ для всіх α, β , і $u_{\alpha, \beta}$ представляє ймовірність того, що індивід $\beta \in \Omega$ мутує в індивід $\alpha \in \Omega$. Тоді $(\hat{U} \cdot \vec{p})_\alpha$ є ймовірністю появи індивіда типу α після застосування мутації для популяції \vec{p} .

Оператор схрещування визначається як $C: \Lambda \rightarrow \Lambda$, де

$$C(\vec{p}) = (\vec{p}^t \cdot \hat{C}_1 \cdot \vec{p}, \dots, \vec{p}^t \cdot \hat{C}_N \cdot \vec{p})$$

і $\hat{C}_1, \dots, \hat{C}_N$ – послідовність симетричних матриць розміру $N \times N$ з дійсними матричними елементами. Тут $\hat{C}_\alpha(\vec{p}) = \vec{p}^t \cdot \hat{C}_\alpha \cdot \vec{p}$ представляє ймовірність того, що індивід α генерується шляхом застосування схрещування до популяції \vec{p} .

Поєднуючи відображення селекції, мутації та кросоверу, отримаємо

повний оператор G для генетичного алгоритму

$$\hat{G}: \Lambda \rightarrow \Lambda, \quad \hat{G}(\vec{p}) = \hat{C} \circ \hat{U} \circ F(\vec{p}). \quad (1)$$

Якщо евристична функція $G_\alpha(\vec{p})$ відома, то можна написати матрицю переходу, яка є стохастичною і визначає ймовірність переходу популяції \vec{p} в популяцію \vec{q} :

$$\Theta_{\vec{q}, \vec{p}} = \bar{m}! \prod_{\alpha \in \Omega} \frac{(G_\alpha(\vec{p}))^{\bar{m}q_\alpha}}{(\bar{m}q_\alpha)!} \quad (2)$$

де $G_\alpha(\vec{p})$ – ймовірність утворення індивіда α в наступному поколінні, $\bar{m}q_\alpha$ – кількість копій індивіда α в популяції \vec{q} , \bar{m} – число індивідів в популяції.

Старші власні значення матриці переходу визначають властивості збіжності простого генетичного алгоритму – в літературі було показано, що швидкість збіжності визначається другим за величиною власним значенням після найстаршого власного значення.

У другому розділі «Метод нецентрованих головних компонент як основа для нового оператора в генетичному алгоритмі для GeantV» присвячений розробленню методу нецентрованих головних компонент (НГК), який теоретично обґрунтовує введення нового оператора в генетичний алгоритм для оптимізації GeantV. Тут запропоновано математичну модель методу нецентрованих головних компонент, яка теоретично обґрунтовує введення нового оператора в генетичний алгоритм для оптимізації GeantV, та отримано формулу для оцінки наближень при застосуванні методу для аналізу даних, яка характеризується тим, що пришвидчує перевірку чутливості параметрів та забезпечує контролювання збереження максимуму інформації при цих наближеннях.

Для оптимізації роботи пакету GeantV при обробці експериментальних даних ідентифіковано параметри оптимізації, які є важливими для проведення симуляцій транспорту елементарних частинок (наприклад, розмірність вектору, що зберігає інформацію про характеристики частинок тощо). Це дало змогу визначити матрицю даних розміром $m \times n$

$$\hat{X} = \{X_{\alpha,i}\} = \{(\vec{x}_\alpha)_i\} = \{\vec{x}_\alpha\}, \quad (3)$$

яка містить значення вимірювань цих параметрів. В цій матриці індекс i нумерує параметри оптимізації ($i = 1, \dots, n$) та індекс α нумерує кількість вимірювань параметрів $\alpha = 1, \dots, m$ (для m вимірювань).

В рамках формулювання генетичного алгоритму матриця даних

описується через m -вибірок даних з n -вимірному простору. У цьому випадку m – кількість індивідів в популяції, n – розмір індивіда (n – розмірність вектору генів $\vec{x} = \{\vec{x}_i\}$) і вектор популяції конструюється як $(\vec{x}_1, \vec{x}_2, \dots, \vec{x}_m) \dots$

Застосування методу головних компонент (МГК) дозволяє перевірити чутливість до негативного впливу «шуму» в значеннях параметрів в генетичному алгоритмі та можливу кореляцію між ними. Для цього ми вводимо оператор, що заснований на застосуванні прямого та зворотнього МГК, для очищення даних від «шуму» в процедурі оптимізації параметрів пакету GeantV за допомогою генетичного алгоритму. Порівнюючи класичну та «нецентровану» версії МГК, можна показати, що метод нецентрованих головних компонент (МНГК) виявляється ефективнішим для очищення даних від шумів.

Головним об'єктом в МНГК є матриця нецентральної моментів

$$\hat{T} = \frac{1}{m} \hat{X}^t \cdot \hat{X}. \quad (4)$$

Ця матриця має ортонормовані власні вектори \vec{w}_j з відповідними власними значеннями t_j

$$\hat{T} \cdot \vec{w}_j = t_j \cdot \vec{w}_j, \quad \vec{w}_i^t \cdot \vec{w}_j = \delta_{i,j}, \quad 1 \leq i, j \leq n. \quad (5)$$

З власних векторів сконструюємо матрицю $W_{i,j} = \{\vec{w}_j\} = (w_i)_j$, яка діагоналізує матрицю \hat{T} :

$$\hat{W}^t \cdot \hat{T} \cdot \hat{W} = \hat{\Delta}, \quad \Delta_{i,j} = t_i \delta_{i,j}, \quad \hat{W}^t \cdot \hat{W} = \hat{I}. \quad (6)$$

Власний вектор \vec{w}_j визначає j -ту «нецентровану» головну компоненту $\vec{\theta}_j = \{(\theta_j)_\alpha\} = \{\vec{x}_\alpha \cdot \vec{w}_j\}$ ($\alpha = 1, \dots, m$). Визначимо матрицю нецентрованих головних компонент $\Theta_{\alpha,j} = \{\vec{\theta}_j\} = \{(\theta_j)_\alpha\}$, яку можна виразити через матрицю даних:

$$\Theta_{\alpha,j} = X_{\alpha,i} W_{i,j}, \quad \Theta_{i,\alpha}^t \Theta_{\alpha,j} = m \Delta_{i,j} = m t_i \delta_{i,j}, \quad 1 \leq \alpha \leq m. \quad (7)$$

У випадку МНГК зручно використовувати наближення, що контролюється малістю так званого параметра власних значень. Ідея полягає у тому аби використовувати МНГК для сингулярного представлення (SVD) нецентрованої матриці даних.

Для зручності визначимо матрицю $\tilde{\Theta}_{\alpha,j}$ як

$$\Theta_{\alpha,j} = \sqrt{m} \tilde{\Theta}_{\alpha,i} \Delta_{i,j}^{1/2}, \quad \Delta_{i,j}^{1/2} = t_i^{1/2} \delta_{i,j}, \quad \tilde{\Theta}_{i,\alpha}^t \tilde{\Theta}_{\alpha,j} \delta_{i,j}. \quad (8)$$

За допомогою цієї матриці нескладно отримати сингулярне представлення для нецентрованої матриці даних

$$X_{\alpha,i} = \sqrt{m} \tilde{\Theta}_{\alpha,k} \Delta_{k,j}^{1/2} W_{j,i}^t. \quad (9)$$

Коли матриця нецентрованих других моментів \hat{T} має $(n-q)$ найменших власних значень $t_j = 1, q+1 \leq j \leq n$, то можна використати наближення, що контролюється малістю параметра власних значень, і отримати наближену матрицю даних $\tilde{X}_{\alpha,i}$ рангу q

$$\tilde{X}_{\alpha,i} = \sqrt{m} \tilde{\Theta}_{\alpha,k} \tilde{\Delta}_{k,j}^{1/2} W_{j,i}^t = \sqrt{m} \left(t_1^{1/2} \tilde{\Theta}_{\alpha,1} W_{1,i}^t + \dots + t_q^{1/2} \tilde{\Theta}_{\alpha,q} W_{q,i}^t \right), \quad (10)$$

де матриця власних значень $\tilde{\Delta}_{k,j}$ має ранг q ($t_{q+1} = t_{q+2} = \dots = t_n = 0$). Тоді можна наблизити матрицю $X_{\alpha,i}$ з рангом n за допомогою матриці $\tilde{X}_{\alpha,i}$ з рангом q .

Нескладно оцінити середньоквадратичну похибку η_q для наближення

$$\eta_q = \frac{1}{mn} \sum_{\alpha=1}^m \sum_{i=1}^n (X_{\alpha,i} - \tilde{X}_{\alpha,i})^2 = \frac{1}{n} \sum_{k=q+1}^n t_k.$$

Похибка є мінімальною, коли матриця нецентральної других моментів \hat{T} має $(n-q)$ найменших власних значень $t_j = 1, q+1 \leq j \leq n$.

В дисертації на основі цих результатів запропоновано метод покращення збіжності генетичного алгоритму через доповнення множини стандартних операторів генетичного алгоритму (селекції, мутації та схрещування) новим оператором P , який реалізує МНГК на просторі популяцій. Тоді новий генетичний алгоритм $G(\vec{p}) = P \circ U \circ F(\vec{p})$ пришвидшує перевірку чутливості параметрів пакета GeantV до оптимізації за допомогою генетичного алгоритму.

У третьому розділі, який називається "*Результати застосування НГК-оператора для еталонних тестів багатокритеріальних задач*", приведені результати тестування нового НГК-оператора при застосуванні його в стандартних тестових багатокритеріальних задачах для генетичних алгоритмів. Тут запропоновано метод прискорення збіжності генетичного алгоритму до оптимального Парето фронту для задач багатокритеріальної оптимізації, який характеризується включенням в генетичний алгоритм нового генетичного оператора (НГК-оператора), що побудований на основі методу нецентрованих головних компонент. В розділі приведені результати тестування нового НГК-оператора при застосуванні його в

стандартних тестових багатокритеріальних задачах для генетичних алгоритмів.

Особливістю еталонних тестів DTLZ (Deb, K., Thiele, L., Laumanns, M., Zitzler, E. (2001)) для багатокритеріальних задач є те, що число цільових функцій в цих багато-цільових проблемах масштабуються до будь-якої кількості цілей. Тестові задачі DTLZ дають змогу досліджувати властивості багатокритеріальних задач контрольованим чином, оскільки для них відомі характеристики оптимального фронту Парето.

У цьому розділі зроблено порівняння результатів застосування алгоритмів NSGA-II [Деб, К., 2002] і NSGA-III [Деб, К., 2012] з включенням НГК-оператора для тестів DTLZ. Зауважимо, що ефективність алгоритму NSGA-III перевищує NSGA-II, але результати NSGA-II представлено як доказ ефективності нашого підходу. Використовуючи тести DTLZ, в розділі був проведений порівняльний аналіз ефективності різних способів включення НГК-оператора в генетичний алгоритм за допомогою застосування оцінки середнє квадратичної похибки й показано, що метод, зображений на рисунку 1, є найбільш ефективним для збіжності модифікованого генетичного алгоритму до фронту Парето.

На рисунках 4 і 5 представлено розподіл параметрів (середнє та стандартне значення відхилення (дисперсію)) та поведінку значення вартості залежно від використовуваних алгоритмів з включенням НГК-оператора.

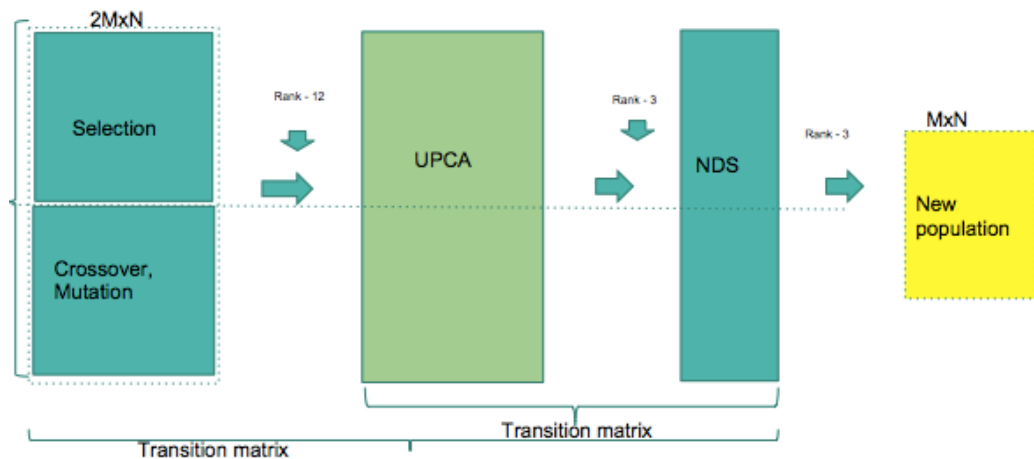


Рисунок 1 – Обрана схема генетичного алгоритму

Порівнюючи рисунки 4, 5, де було застосовано процедуру видалення шуму за допомогою НГК-оператора, з рисунками 2, 3, де НГК-оператор не був застосований, можна спостерігати швидку збіжність до ідеальних значень параметрів у першому випадку.

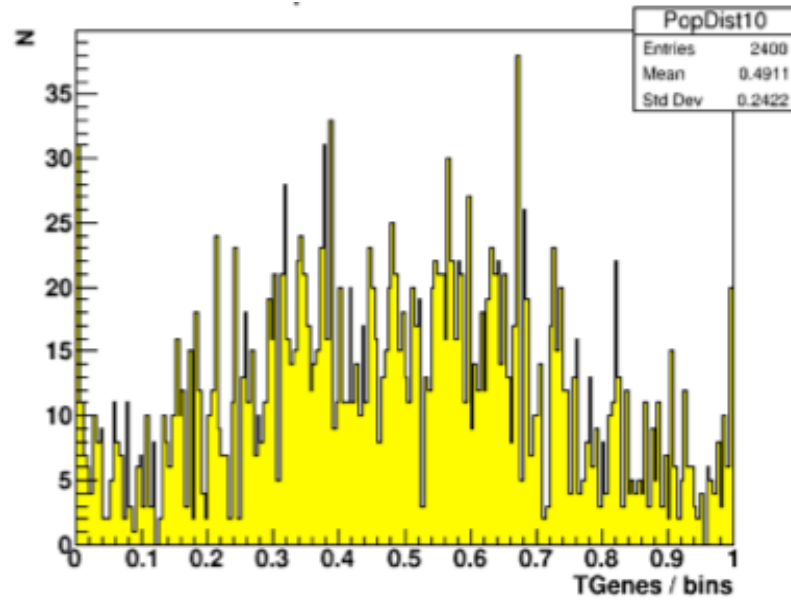


Рисунок 2 – Розподіл популяцій на 10-тій генерації NSGA-II для DTLZ2

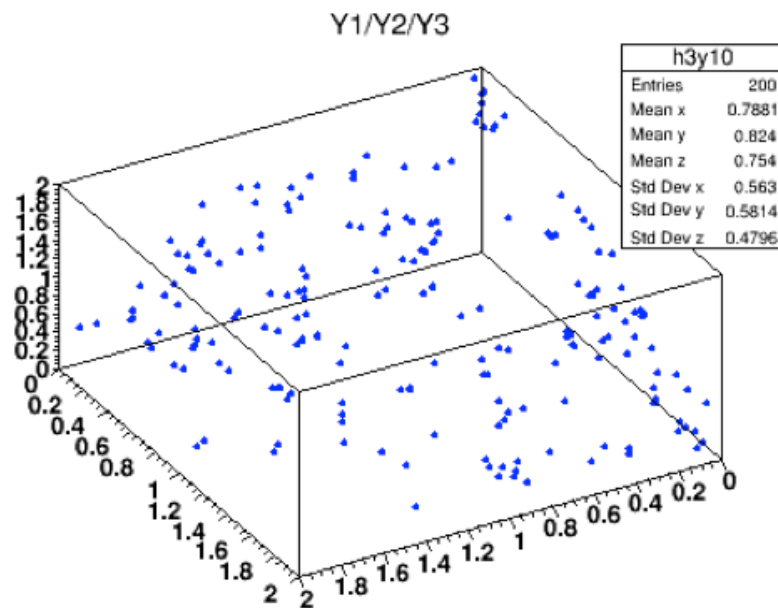


Рисунок 3 – Фронт Парето на 10-тій генерації NSGA-II для DTLZ2

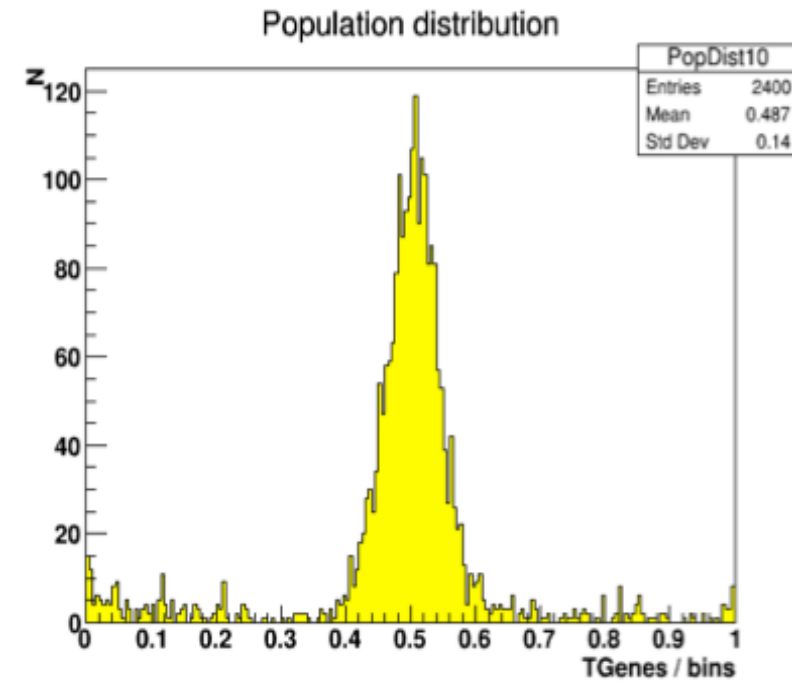


Рисунок 4 – Розподіл популяцій на 10-тій генерації NSGA-II для DTLZ2 із застосуванням НГК-оператора

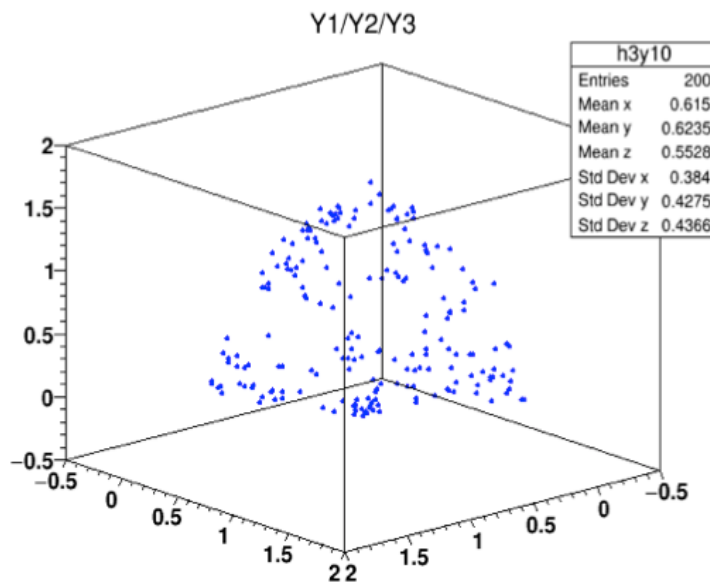


Рисунок 5 – Фронт Парето на 10-тій генерації NSGA-II для DTLZ2 із застосуванням НГК-оператора

У четвертому розділі «Тестування застосування НГК-оператора в генетичному алгоритмі, що використовується для оптимізації продуктивності пакету симуляцій транспорту частинок GeantV» розглядаються результати застосування НГК-оператора до оптимізації продуктивності пакету симуляцій транспорту частинок GeantV та сформульовані рекомендації для застосування запропонованої в дисертаційній роботі модифікації генетичних алгоритмів для оптимізації пакетів симуляцій транспорту частинок, які перевірені при застосуванні НГК-оператора до оптимізації продуктивності пакету GeantV. Використання НГК-оператора дало змогу збільшити продуктивність використання обчислювальних ресурсів для аналізу даних у фізиці високих енергій.

Найважливішим критерієм ефективності паралельного програмування є швидкодія програми, її продуктивність. На продуктивність впливають різні фактори (архітектура обчислювальної системи, методи управління ресурсами, ефективність паралельного алгоритму, особливості структури даних і т. д). Починаючи з деякого значення, збільшення числа процесорів дає тільки невеликий приріст продуктивності. Залежність коефіцієнта прискорення від числа процесорів і ступеня паралелізму алгоритму (відносної частки паралельної частини) носить назву закону Амдала.

Очікується, що це покращення пропускнуєї спроможності обробки подій частково компенсує необхідність збільшення часу та ресурсів для симуляції подій в наступних експериментах на ВАК при вищих енергіях зіткнення частинок на прискоричувах з народженням великої кількості вторинних частинок. Пакет GeantV отримує значні переваги від використання своїх оптимізованих компонент, наприклад, таких як нова бібліотека моделювання геометрії, яка забезпечує кілька нових функцій, включаючи векторизацію та прозорий одночасний доступ до декількох комп'ютерних архітектур.

Перед дослідженням програмного пакету GeantV, важливо оцінити позитивні та слабкі сторони його попередника Geant4. Geant4 - це програмний пакет для моделювання проходження частинок через речовину. Сфери його застосування включають фізику та астрофізику високих енергій, фізику прискорювачів елементарних частинок та ядерну фізику, а також дослідження в медичній та космічній науках. Для ознайомлення з специфікою програмного пакету Geant4 дисертант приймав участь в обробці даних з народження та розпаду чарівних адронів з фіксованою мішенню в експериментах SHiP.

Завдяки зусиллям команди Geant4, яка прагне вдосконалити обчислювальну продуктивність пакету, код Geant4 вже має високий рівень оптимізації, що ускладнює отримання додаткового підвищення

продуктивності у майбутньому. Через його складний дизайн, що має глибокі ієрархічні структури класів та стеків викликів, Geant4 важко експлуатувати на кращих сучасних комп'ютерних архітектурах, що мають багатоядерну структуру процесорів.

Цей факт був одним з важливих для мотивації проведення досліджень в рамках цієї кандидатської дисертації, оскільки ми намагалися знайти нові інноваційні алгоритми для підвищення ефективності поточного стану симуляцій подій при високих енергіях.

Для моделі оптимізації продуктивності GeantV можемо виділити вісім рівнів паралелізму, де дрібноблочний паралелізм є єдиним методом, що надає можливості для стійкого зростання пропускної здатності, і зменшення часу опрацювання даних. Зокрема, багатоядерний паралелізм дає змогу зменшити обсяг пам'яті та час опрацювання даних, але не пропускну здатність, тоді як багатопроцесовий рівень опрацювання дає можливість виконувати велику кількість паралельних завдань. Всі ці опції опрацювання даних доступні для GeantV із збільшуванням розмірності простору параметрів, які необхідно оптимізувати для багатокритеріальної задачі оптимізації програмного пакету GeantV.

Для налагодження роботи GeantV важливе конфігурування інфраструктури керування подіями, що працює на нодах та забезпечуються матрицею налаштованих параметрів – це загальна кількість подій, кількість буферизованих подій, поріг для визначення пріоритетів подій, кількість потоків процесора, політика планування розміщення NUMA локальності, вихідний розмір вектору, порогове значення перемикання між скалярним та векторним режимами, спеціальний поріг під час активації кошика в диспетчері завдань, поріг пам'яті, поріг знищення подій та кількість пропагаторів. З іншого боку необхідна інформація про характеристики нодів, які можна оцінити за допомогою агрегованої матриці функцій пристосованості продуктивності – це оптимізація загального часу виконання симуляції, середній час оптимізації для створення однієї популяції, зменшення використання пам'яті додатків, оптимізація кількості викликів інструкцій для моделювання, специфічне для поточної архітектури зниження спорідненості пам'яті (NUMA оптимізація), споживаної мікропроцесором.

Подальше вдосконалення для інфраструктури керування подіями було досягнуто на основі еволюційного алгоритму параметричного налагодження конфігурацій моделювання та параметрів диспетчеризації подій.

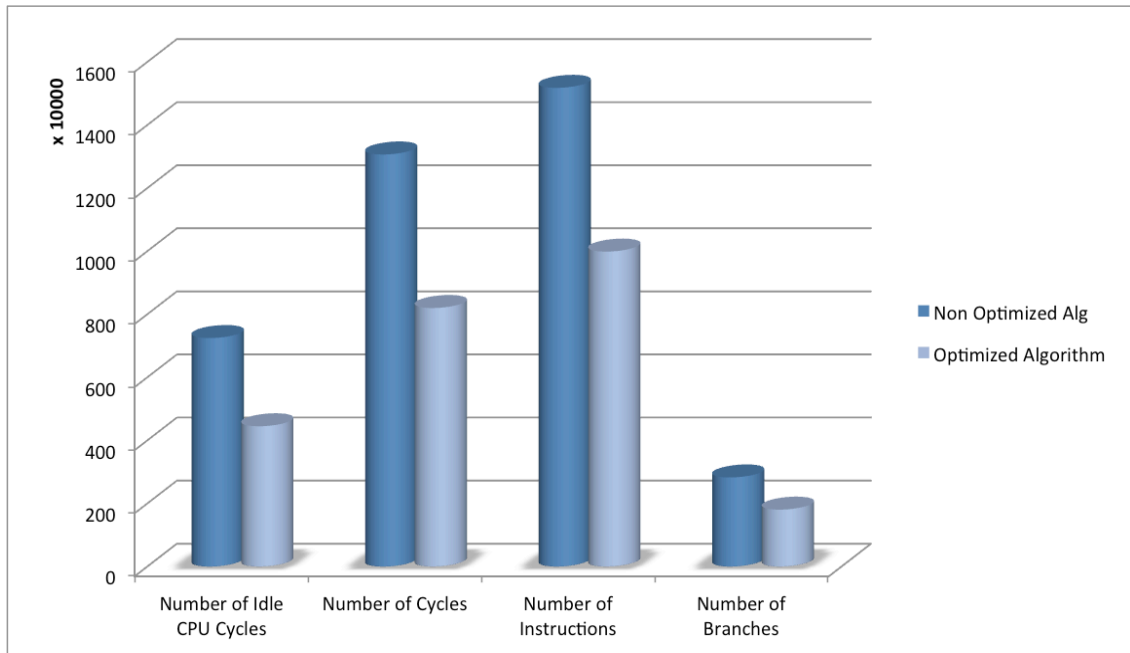


Рис. 6 – Оптимізація продуктивності для функцій пристосованості, спеціалізованих для оцінки продуктивності додатку для процесорів Intel(R)Core i7

У розділі зроблено порівняння результатів застосування алгоритму з включенням НГК-оператора для пакету симуляцій транспорту частинок для аналізу простого детектора в експерименті CMS. На рисунках 6, 7 і 8 представлено результати симуляцій оптимізації продуктивності пакету GeantV для процесорів Intel(R) Ivy Bridge та Intel(R) Core i7.

Зокрема, результати застосування процедури зменшення шуму за допомогою НГК-оператора для оптимізації роботи на архітектурі Intel Core i7 представлено на рисунку 6, на яких бачимо швидкий прогрес у зменшенні загального часу роботи для групи запущених задач з процедурою генетичної оптимізації. А результати застосування НГК-оператора в генетичному алгоритмі для процесорів Intel(R) Ivy Bridge представлено на рисунках 7, 8, на яких видно швидку збіжність до значень, що можуть бути наближеними до ідеальних значень параметрів вже після 15-ти генерацій генетичного алгоритму.

Результати показують, що інтеграція генетичного алгоритму в GeantV для оптимізації продуктивності симуляцій є одним з важливих кроків до модернізації програмових пакетів, використовуваних у фізиці високих енергій.

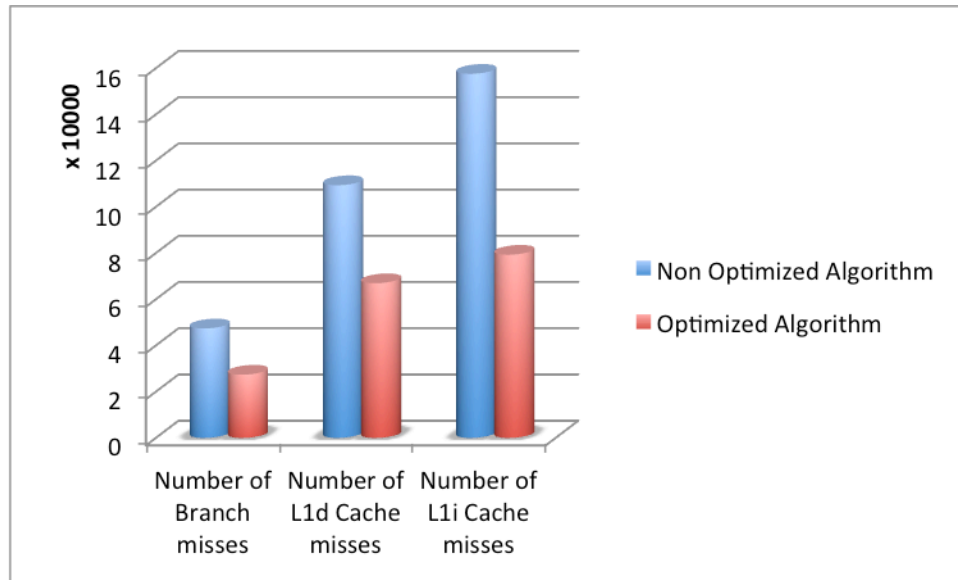


Рисунок 7 – Оптимізація продуктивності для функцій пристосованості, для оцінки продуктивності додатку для процесорів Intel(R) Ivy Bridge

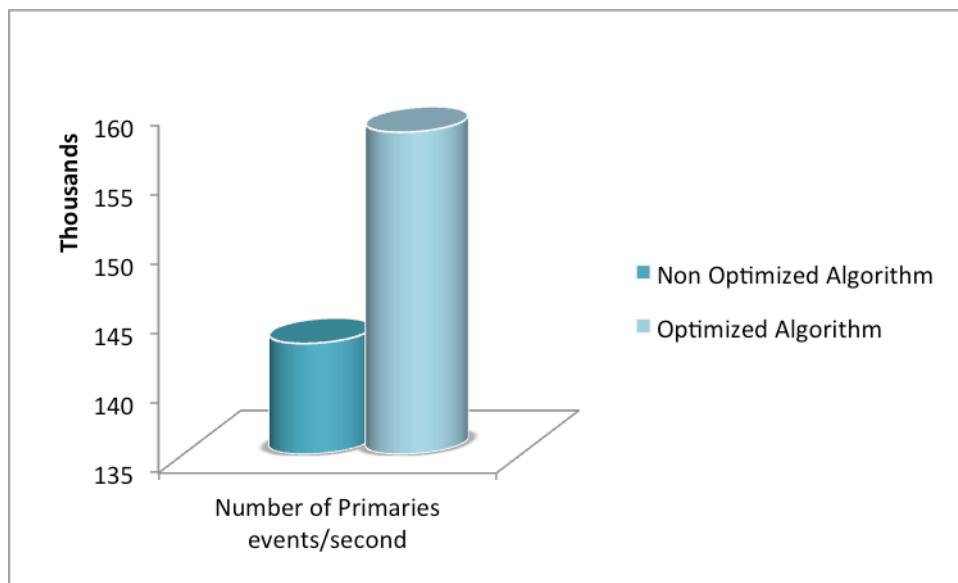


Рисунок 8 – Кількість зіткнень первинних частинок за одиницю часу для процесорів Intel(R) Ivy Bridge

У **Висновку** сформульовано основні результати дисертаційної роботи, що виносяться на захист.

Науково-технологічні дослідження, що проведені в рамках дисертаційної роботи, характеризуються успішним вирішенням актуальної науково-практичної задачі розробки та застосування стохастичної

оптимізації продуктивності САЕ-системи GeantV в фізиці високих енергій на основі модифікації генетичних алгоритмів за допомогою НГК-оператора.

1. Сформульовано багатоцільову задачу оптимізації продуктивності САЕ-системи автоматизованого проектування GeantV, обґрунтовано вибір генетичного алгоритму в якості базової процедури такої оптимізації, визначено простір розв'язків та цільовий простір. Використовуючи цей аналіз, представлення генетичного алгоритму як динамічної системи через його зв'язок з марковськими ланцюжками та аналіз властивостей власних значень матриці переходу, вдалося визначити оцінки фронту Парето та оптимальної точки Парето.

2. Вперше розроблено метод конвергенції генетичних алгоритмів, який відрізняється від існуючих розробкою нового математичного алгоритму за допомогою включення в генетичний алгоритм нового генетичного оператора (НГК-оператора) що побудований на основі методу нецентрованих головних компонент та дозволяє прискорити збіжність генетичного алгоритму до оптимального Парето фронту для задач багатокритеріальної оптимізації. Отримані формули для оцінки наближень при аналізі даних за допомогою цього методу, які забезпечують контролювання збереження максимуму інформації при цих наближеннях. Запропоновано підходи що до включення цього методу в генетичний алгоритм через введення нового оператора (НГК-оператора) в динамічну систему генетичного алгоритму.

3. Розроблена програмна реалізація включення в генетичний алгоритм нового оператора на основі методу нецентрованих головних компонент (НГК-оператор) і показано, що його застосування в тестових багатоцільових задачах оптимізації забезпечує прискорення збіжності генетичного алгоритму до оптимального фронту Парето.

4. Удосконалено технологію інтеграції методу конвергенції генетичних алгоритмів для стохастичної оптимізації продуктивності САЕ-системи моделювання транспорту частинок GeantV, за допомогою імплементації цих алгоритмів в спеціальній бібліотеці для типової САЕ-системи автоматизованого проектування в фізиці високих енергій, а саме програмного пакету GeantV і проведено дослідження його продуктивності в порівнянні із застосуванням немодифікованого генетичного алгоритму. Дослідження показали, що для еталонних тестів багатоцільових задач оптимізації загальний час виконання партій симуляцій за допомогою модифікованого генетичного алгоритму скоротився приблизно на 18 % (для процесорів Intel(R) Core i7) і на 27% з одночасним зменшенням часу використання CPU на ~34% (для процесорів Intel(R) Xeon(R) E5-2695).

Результати, отримані в ході цього дослідження, є доказом ефективності концепції оптимізації продуктивності САЕ-системи GeantV з використанням модифікованих генетичних алгоритмів для стохастичної

оптимізації. Це метод можна застосовувати для оптимального встановлення GeantV-додатків на суперкомп'ютерах або обчислюваних кластерах, одночасно з можливістю налаштування високопаралельних задач у неоднорідному обчислюваному середовищі, для досягнення масштабованості для високопродуктивних обчислювань.

ОСНОВНІ ПРАЦІ, ОПУБЛІКОВАНІ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

Статті у наукових фахових виданнях

1. Shadura O. Multivariate convergence-targeted operator for the genetic algorithm / O. Shadura, A. Petrenko, S. Svistunov // Системні дослідження та інформаційні технології. – 2017. – № 1. – С. 126–140. – Бібліогр.: 17 назв. – *Журнал включено в міжнародні наукометричні бази: Index Copernicus, DOAJ, EBSCOhost Reaserch Database, особистий внесок здобувача – розроблено математичний апарат методу нецентрованих головних компонент (НГК).*

2. Оксана Шадура. «Метод головних компонент і оптимізація пакетів фізичного моделювання за допомогою генетичних алгоритмів» // Вісник Університету «Україна», Серія «Інформатика, обчислювальна техніка та кібернетика». – 2019. – №1 (22) – С. 198–209. – Бібліогр.: 16 назв. – *Особистий внесок здобувача – проведений аналіз вибору оптимальної модифікації генетичного алгоритму за рахунок включення в алгоритм НГК-оператора та досліджена продуктивність такого генетичного алгоритму.*

3. Оксана Шадура. Модифікація генетичних алгоритмів на основі методу нецетрованих головних компонент та стандартні тести, // World Science – 2019. – № 4(44). – С. 4–11. – Бібліогр.: 10 назв. – DOI: 10.31435/rsglobal_ws. – *Журнал включено в міжнародні наукометричні бази: Index Copernicus, Google Scholar, academia.edu. Особистий внесок здобувача – зформульовано підхід до оптимізації програмного пакету GeantV.*

4. Oksana Shadura. Performance Optimization of Physics Simulations Through Genetic Algorithms. / Oksana Shadura, Federico Carminati and Anatoliy Petrenko. // Journal of Computer Science. – 2019. – Volume 15. – Issue 1. – P. 57–66. – DOI 10.3844/jcssp.2019.57.66. – *Журнал включено в міжнародні наукометричні бази: Scopus, EBSCO, особистий внесок здобувача – було досліджено вибір ключових параметрів програмного пакету GeantV для оптимізації продуктивності цього пакету та перевірено обраних параметрів.*

5. Оксана Шадура. Оптимізація пакету фізичного моделювання GeantV за допомогою генетичних алгоритмів. // Вчені записки Таврійського національного університету імені В. І. Вернадського, Серія: Технічні науки – 2019. – Том 30 (69). – № 3. – С. 206–213. *Журнал включено в міжнародні наукометричні бази: Index Copernicus, особистий внесок здобувача – було показано, що модифікований генетичний алгоритм з включенням НГК-оператора прискорює оптимізацію продуктивності програмного пакету GeantV при симуляції транспорту частинок через простий калориметр та збільшує ефективність використання обчислювальних ресурсів для аналізу даних у фізиці високих енергій.*

Тези доповідей в збірках матеріалів конференцій

1. O. Shadura et al. The experimental facility for the Search for Hidden Particles at the CERN SPS / Ahdida C., Albanese R. et al. // Journal of Instrumentation (JINST). – 2019. – Volume 14. – Issue 3. – P. 3025–3046 – DOI 10.1088/1748-0221/14/03/p03025. – *Особистий внесок здобувача – було проведено аналіз слабких та сильних сторін програмного пакету Geant4, який є попередником пакету нового покоління.*

2. Oksana Shadura. Stochastic performance tuning of complex simulation applications using unsupervised machine learning. // Oksana Shadura, Federico Carminati // 2016 IEEE Symposium Series on Computational Intelligence – 2016. – ISBN 978-1-5090-4240-1. – *Особистий внесок здобувача – розроблено підхід до оптимізації програмного пакету GeantV.*

3. G. Amadio and all. Stochastic optimization of GeantV code by use of genetic algorithms // G. Amadio and J. Apostolakis and M. Bandieramonte and S.P. Behera and R. Brun and P. Canal and F. Carminati and G. Cosmo and L. Duhem and D. Elvira and G. Folger and A. Gheata and M. Gheata and I. Goulas and F. Hariri and S.Y. Jun and D. Konstantinov and H. Kumawat and V. Ivantchenko and G. Lima and T. Nikitina and M. Novak and W. Pokorski and A. Ribon and R. Seghal and O. Shadura and S.Vallecora and S. Wenzel, Stochastic optimization of GeantV code by use of genetic algorithms // Journal of Physics: Conf. Ser. Offline Computing – 2017. – Vol. 898, N. 4, – P. 042026 – IOP Publishing. – *Особистий внесок здобувача – розроблена імплементація системи оптимізації програмного пакету GeantV.*

4. Amadio G and all. GeantV: from CPU to accelerators // G. Amadio, A. Ananya, J. Apostolakis, A. Arora, M. Bandieramonte, A. Bhattacharyya, C. Bianchini, R. Brun, P. Canal, F. Carminati, L. Duhem, D. Elvira, A. Gheata, M. Gheata, I. Goulas, R. Iope, S. Jun, G. Lima, A. Mohanty, T. Nikitina, M. Novak, W. Pokorski, A. Ribon, R. Seghal, O. Shadura, S. Vallecora, S. Wenzel, Y. Zhang // Journal of Physics: Conf. Ser. Computing Technology for

Physics Research – 2016. – Vol. 762, N. 1. – P. 012019. – IOP Publishing. – *Особистий внесок здобувача – опис термінології пакету GeantV, що визначає основну проблематику та можливі алгоритми, що мають бути дослідженні для вирішення проблеми оптимізації.*

5. Amadio G and all. The GeantV project: preparing the future of simulation // G. Amadio, J. Apostolakis, M. Bandieramonte, A. Bhattacharyya, C. Bianchini, R. Brun, P. Canal, F. Carminati, J. C. De Fine Licht, L. Duhem, V. D. Elvira, A. Gheata, R. L. Iope, G. Lima, A. Mohanty, T. Nikitina, M. Novak, W. Pokorsky, R. Sehgal, O. Shadura, S. Vallecorsa, S. Wenzel // Journal of Physics: Conf. Ser. Offline software – 2015. – Vol. 664, N. 7. – P. 072006 – IOP Publishing. – *Особистий внесок здобувача – робота над системою паралелізму у пакеті GeantV та інтеграцією системи оптимізації в пакет GeantV.*

6. Apostolakis J and all. Adaptive track scheduling to optimize concurrency and vectorization in GeantV // J. Apostolakis, M. Bandieramonte, G. Bitzes, R. Brun, P. Canal, F. Carminati, J. C. De Fine Licht, L. Duhem, V. D. Elvira, A. Gheata, S. Y. Jun, G. Lima, M. Novak, R. Sehgal, O. Shadura, S. Wenzel // Journal of Physics: Conf. Ser. Advanced Computing and Analysis Techniques in Physics Research – 2015. – Vol. 608, N. 1. – P. 012003 – IOP Publishing. – *Особистий внесок здобувача – визначення основної проблематики для оптимізації пакету GeantV та параметрів, що мають бути дослідженні для оптимізації пакету GeantV.*

7. J. Apostolakis and all. Towards a high performance geometry library for particle–detector simulations // J. Apostolakis and M. Bandieramonte and G. Bitzes and R. Brun and P. Canal and F. Carminati and G. Cosmo and J. C. De Fine Licht and L. Duhem and V. D. Elvira and A. Gheata and S. Y. Jun and G. Lima and T. Nikitina and M. Novak and R. Sehgal and O. Shadura and S. Wenzel // Journal of Physics: Conf. Ser. Computing Technology for Physics Research – 2015. – Vol. 608, N. 1, – P. 012023 – IOP Publishing. – *Особистий внесок здобувача – визначення основної проблематики для оптимізації пакету VecGeom та можливих алгоритмів та параметрів, що мають бути дослідженні для вирішення проблеми оптимізації параметрів пакету GeantV.*

АНОТАЦІЯ

Шадура О.В. Стохастична оптимізація продуктивності пакету симуляції транспорту частинок в фізиці високих енергій. – Рукопис.

Дисертація на здобуття наукового ступеня кандидата технічних наук за спеціальністю 05.13.12 – Системи автоматизації проектувальних робіт. – Національний технічний університет

України «Київський політехнічний інститут імені Ігоря Сікорського».
– Київ, 2019.

В дисертаційній роботі запропоновано новий підхід для дослідження оптимізації продуктивності САЕ-системи моделювання транспорту частинок у фізиці високих енергій GeantV на основі стохастичних методів. Задачу оптимізації сформульовано за допомогою генетичного алгоритму, представленого у вигляді динамічної системи завдяки його зв'язку з марковськими ланцюжками. Розроблено метод нецентрованих головних компонент, його математичні основи та отримано формули для оцінки наближень при його застосуванні для аналізу даних. Проведено порівняльний аналіз використання методу головних компонент при його застосуванні для зцентрованих і нецентрованих даних та знайдено еквівалентність результатів використання обох представлень даних на основі методу оцінювання середньоквадратичної похибки. Запропоновано новий генетичний оператор, побудований на основі методу нецентрованих головних компонент (НГК-оператор), що дає змогу прискорити збіжність генетичного алгоритму до оптимального фронту Парето для задач багатокритеріальної оптимізації. Знайдено оптимальну інтеграцію НГК-оператора в типовий генетичний алгоритм. Досліджено ефективність при його застосуванні в еталонних тестах багатоцільових задач та для оптимізації продуктивності САЕ-системи моделювання транспорту частинок GeantV у порівнянні з неоптимізованою версією такого самого програмового пакету. Встановлено, що середній приріст продуктивності досягає до 20 % над неоптимізованим запуском пакету у неоднорідному обчислювальному середовищі.

Ключові слова: генетичний алгоритм, динамічна система, нерухомі точки, стохастична оптимізація, генетичний оператор, аналіз основних компонентів, оптимізація продуктивності.

АННОТАЦИЯ

Шадур О.В. Стохастическая оптимизация производительности пакета симуляции транспорта частиц в физике высоких энергий. – Рукопись.

Диссертация на соискание ученой степени кандидата технических наук по специальности 05.13.12 – Системы автоматизации проектировочных работ. – Национальный технических университет Украины «Киевский политехнический

институт имени Игоря Сикорского». – Киев, 2019.

В диссертационной работе предложен новый подход для исследования оптимизации производительности САЕ-системы моделирования транспорта частиц в физике высоких энергий GeantV на основе стохастических методов. Задачу оптимизации сформулировано с помощью генетического алгоритма, представленного в виде динамической системы благодаря его связи с марковскими цепочками. Разработан метод нецентрированных главных компонент, его математические основы и получены формулы для оценки приближений при его применении для анализа данных. Проведен сравнительный анализ использования метода главных компонент при его применении для центрированных и нецентрированных данных, и найдена эквивалентность результатов использования обоих представлений данных на основе метода оценки среднеквадратической погрешности. Предложен новый генетический оператор, построенный на основе метода нецентрированных главных компонент (НГК-оператор), который дает возможность ускорить сходимость генетического алгоритма к оптимальному фронту Парето для задач многокритериальной оптимизации. Найдена оптимальная интеграция НГК-оператора в типовой генетический алгоритм и исследованы его эффективность при применении для эталонных тестов многоцелевых задач и для оптимизации производительности САЕ-системы моделирования транспорта частиц GeantV по сравнению с неоптимизированной версией того же программного пакета. Средний прирост производительности достигает до 20% выше неоптимизированного запуска пакета в неоднородной вычислительной среде.

Ключевые слова: Генетический алгоритм, динамическая система, фиксированные точки, стохастическая оптимизация, генетический оператор, анализ основных компонентов, оптимизация производительности.

SUMMARY

Shadura O.V. Stochastic optimization of the performance of a particle transport simulation package in High Energy Physics – Manuscript.

Dissertation for a Degree of Candidate of Technical Sciences (PhD), specialization 05.13.12 – «Automation of systems for computer aided

design». – National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute». – Kyiv, 2019.

The dissertation is dedicated to study the optimization strategies using genetic algorithms to achieve the best computing performance of the particle transport simulation package GeantV for the High Energy Physics simulations. It had been developed to solve the problems of processing the large data sets in experiments at the Large Hadron Collider (LHC). In the scope of this dissertation had been developed a software library integrating the genetic algorithms as a part of the GeantV software package. It will be used to optimize the performance of GeantV software package for the particle transport simulation, when used for processing the physics data, collected by experiments on the LHC. The mathematical model of the non-centered principal components analysis method is defined together with the formula for estimation of error approximation. The estimation of the error approximation shows that the procedure of reducing the data matrix, based on the selection of eigenvectors of the matrix of non-centered second moments for which it has the smallest eigenvalues, is correct. A modification of the genetic algorithm is defined by introducing into the standard set of genetic algorithm operators (selection, mutation, crossover), a new operator, which is determined by the non-centered principal component analysis method (as a new genetic operator to be used on genetic populations). The results of this study are presenting the proof of the concept of optimization of GeantV performance using evolutionary algorithms with an average gain of 20% of over not optimized run in heterogeneous computing environment. The same method can be used to deploy GeantV applications on supercomputers and clusters for the efficient high-performance computing, while configuring massively parallel GeantV simulations, launched in a non-homogeneous computing environment and providing optimal scalability in high-performance computing environment.

Keywords: genetic algorithm, dynamic system, fixed points, stochastic optimization, genetic operator, principle component analysis, performance optimization.